

Cloud Computing in Data Mining: A Technical Survey

Authors : Ghaemi Reza, Abdollahi Hamid, Dashti Elham

Abstract : Cloud computing poses a diversity of challenges in data mining operation arising out of the dynamic structure of data distribution as against the use of typical database scenarios in conventional architecture. Due to immense number of users seeking data on daily basis, there is a serious security concerns to cloud providers as well as data providers who put their data on the cloud computing environment. Big data analytics use compute intensive data mining algorithms (Hidden markov, MapReduce parallel programming, Mahot Project, Hadoop distributed file system, K-Means and KMediod, Apriori) that require efficient high performance processors to produce timely results. Data mining algorithms to solve or optimize the model parameters. The challenges that operation has to encounter is the successful transactions to be established with the existing virtual machine environment and the databases to be kept under the control. Several factors have led to the distributed data mining from normal or centralized mining. The approach is as a SaaS which uses multi-agent systems for implementing the different tasks of system. There are still some problems of data mining based on cloud computing, including design and selection of data mining algorithms.

Keywords : cloud computing, data mining, computing models, cloud services

Conference Title : ICSET 2014 : International Conference on Software Engineering and Technology

Conference Location : Paris, France

Conference Dates : December 30-31, 2014